통계

문자, 이메일, 뉴스, 광고 등 하루에도 엄청나게 쏟아져 나오는 자료에서
우리에게 유용하고 필요한 정보를 수집하는 능력은 그 어느 때보다 중요하다.
수집된 정보와 자료를 정리, 요약하여 불확실한 미래를 예측하는 것은 현대 사회에서
합리적인 의사 결정을 하는 데 도움이 된다.

배운 내용

- 평균(초등)
- 좌표와 그래프,
  자료의 정리와 해석(중1)

이 단원에서는

Ⅵ-1 대푯값
산포도

Ⅵ-2 산점도와 상관관계

배울 내용

- 확률분포, 통계적 추정
  (확률과 통계)

| 준비 학습 |

1 다음은 학생 5명이 한 학기 동안 자율 동아리 활동을 한 시간을 조사한 것이다. 자율 동아리 활동 시간의 평균을 구하시오. 초등

(단위: 시간)

35	24	23	41	32

2 오른쪽 줄기와 잎 그림은 서준이네 반 학생들이 지난 한 달 동안 받은 전자 우편의 수를 조사하여 그린 것이다. 중1

(1) 서준이네 반의 전체 학생 수를 구하시오.

(2) 전자 우편의 수가 작은 쪽에서 10번째인 전자 우편의 수를 구하시오.

전자 우편의 수 (0|4는 4개)

줄기	잎
0	4 5 6 7 8 9
1	0 1 2 3 4 6 6 8
2	1 1 2 5 7 9 9
3	3 3 4 6 8
4	2 5

대단원 포트폴리오

이 단원을 학습하면서 다음 중에서 하나를 선택하여 작성해 보자.

- ☐ 수학 달력
- ☐ 수학 게임
- ☐ 수학 포스터
- ☐ 수학 만화
- ☐ 수학사 보고서
- ☐ 수학 비주얼 싱킹

● 포트폴리오 예시와 길잡이 ▶ 226쪽

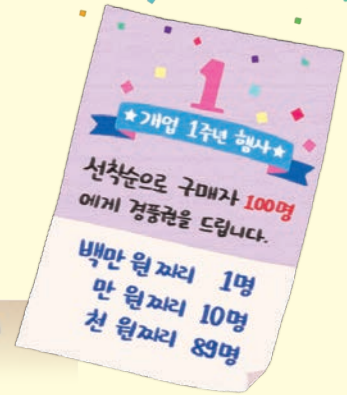
1 대푯값과 산포도

수학 + 사회

시장 경쟁이 치열할수록 상품 광고의 중요성은 커진다. 광고는 소비자에게 상품이나 서비스에 대한 정보를 제공하고 장점을 부각하여 상품에 대한 구매 욕구를 자극한다.

그러나 판매자는 소비자의 강렬한 관심을 불러일으키기 위해 때때로 과장 광고를 하기도 한다. 따라서 소비자는 광고 속 정보가 올바른지 판단해야 한다.

단원 활동 광고 내용을 통해 자료를 대표하는 값을 알아보자.

오른쪽은 어느 가게에서 행사하는 경품권에 대한 광고 내용이야.

활동 1 경품권의 평균 금액과 가장 많이 발행한 경품권의 금액을 각각 구해 보자.

활동 2 위의 광고에서 “구매자에게 백만 원짜리 경품권 증정”이라는 내용이 과장 광고인지 아닌지를 판단하고, 과장 광고라면 **활동 1**에서 구한 두 금액 중에서 하나를 택하여 광고 내용을 쓰고, 그렇게 쓴 이유를 설명해 보자.

위의 활동으로 **알게 된 것**과 **나의 학습 계획**을 적어 보자.

알게 된 것 ▶ 자료의 특징을 나타내는 값을 찾을 수 있다. 예 ☐ 아니요 ☐

학습할 내용 ▶ 대푯값 ▶ 산포도

학습 계획 ______________________

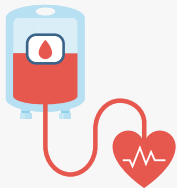
____

학습 계획안 예시
- 예습과 복습을 열심히 하겠다.
- 수업 시간에 집중하겠다.
- 수학에 대한 자신감을 키우겠다.
- 모둠 활동에 적극적으로 참여하겠다.

대푯값

• 중앙값, 최빈값, 평균의 의미를 이해하고, 이를 구할 수 있다.

중앙값은 무엇일까?

개 념 열 기

다음은 어느 헌혈의 집에서 하루 동안 헌혈한 사람 9명의 나이를 조사한 자료이다.

(단위: 세)

58	22	26	24	23	26	20	23	21

1 헌혈한 사람들의 나이의 평균을 구하시오.

2 헌혈한 사람들의 나이를 작은 값부터 크기순으로 나열할 때, 중앙에 위치한 값을 구하시오.

자료를 정리하여 그래프로 나타내면 자료의 분포 상태를 한눈에 알 수 있다. 그러나 자료의 분포 상태의 특성을 알기 위해서는 먼저 그 자료의 중심이 어디에 위치하고 있는지 그 경향을 알아야 한다.

일반적으로 자료 전체의 중심 경향이나 특징을 대표적으로 나타내는 값을 그 자료의 **대푯값**이라고 한다.

대푯값으로 가장 많이 사용하는 것은 평균이다. 평균은 대체로 자료 전체의 중심 경향을 잘 나타내어 주지만 그렇지 않은 경우도 있다.

초등 에서 배웠어요!

$(평균)=\frac{(전체 자료의 합)}{(자료의 개수)}$

위의 개념 열기에서 헌혈한 사람들의 나이의 평균은

$$(평균)=\frac{58+22+26+24+23+26+20+23+21}{9}=27(세)$$

이다.

그런데 9명 중에서 8명의 나이가 평균 27세보다 적고, 1명의 나이가 평균보다 많으므로 평균 27세는 이 자료의 대푯값으로 적절하다고 할 수 없다.

이와 같이 자료의 값 중에서 매우 크거나 매우 작은 값이 있는 경우에는 평균이 이 자료 전체의 중심 경향을 잘 나타낸다고 할 수 없으므로 평균 이외의 다른 대푯값이 필요하다.

자료를 작은 값부터 크기순으로 나열할 때, 자료의 중앙에 위치한 값을 그 자료의 **중앙값**이라고 한다.

중앙값을 구할 때는 자료를 작은 값부터 크기순으로 나열하여 자료의 개수가 홀수이면 중앙에 위치한 하나의 값을 중앙값으로 하고, 자료의 개수가 짝수이면 중앙에 위치한 두 값의 평균을 중앙값으로 한다.

스스로 확인하기

(1) 자료가 9, 7, 10, 8, 24일 때, 이 자료를 작은 값부터 크기순으로 나열하면

7, □, □, 10, □ ← 자료의 개수가 홀수

이므로 중앙값은 □이다.

(2) 자료가 10, 15, 4, 17, 11, 14일 때, 이 자료를 작은 값부터 크기순으로 나열하면

□, 10, 11, □, 15, 17 ← 자료의 개수가 짝수

이므로 중앙값은 $\frac{11+\square}{2}=\square$이다.

다음 자료의 중앙값을 구하시오.

(1) 167, 170, 181, 159, 78, 163, 170

(2) 56, 75, 69, 17, 48, 62, 60, 73

다음은 학생 8명이 제기차기를 한 횟수를 조사한 자료이다. 평균과 중앙값 중에서 이 자료의 대푯값으로 적절한 것을 구하고, 그 이유를 설명하시오.

(단위: 회)

6	5	7	4	6	7	23	6

최빈값은 무엇일까?

다음은 알뜰 나눔 장터에 기증된 티셔츠 10벌의 치수를 조사한 자료이다.

(단위: 호)

95	90	95	90	95
95	100	100	95	105

위의 자료에서 가장 많이 나타난 값은 95호이다. 이와 같이 자료의 값 중에서 가장 많이 나타난 값을 그 자료의 **최빈값**이라고 한다.

최빈값에서 최(最)는 '가장', 빈(頻)은 '자주'라는 뜻이다.

좋아하는 색, 혈액형, 실시간 검색어 등과 같이 자료의 값이 수가 아닌 경우에는 대푯값으로 최빈값을 사용한다.

일반적으로 옷의 치수와 같이 규격화된 자료, 에너지 효율 등급과 같이 등급화된 자료에서는 자료의 대푯값으로서 최빈값이 적절하다.

평균이나 중앙값은 값이 하나로 정해지지만, 최빈값은 자료에 따라서 두 개 이상일 수도 있다.

스스로 확인하기

(1) 자료가 3, 4, 2, 4, 4, 5일 때, 4가 세 번으로 가장 많이 나타나므로 최빈값은 4이다.

(2) 자료가 5, 7, 6, 7, 8, 8일 때, 7과 8이 각각 두 번씩 가장 많이 나타나므로 최빈값은 7과 8이다.

다음 자료의 최빈값을 구하시오.

(1) 55, 57, 58, 55, 55, 57, 53

(2) 16, 19, 21, 17, 18, 16, 22, 17, 20

다음은 어느 신발 가게에서 일주일 동안 판매한 운동화의 크기를 조사한 자료이다. 이 가게에서 가장 많이 준비해야 할 운동화의 크기를 정하려고 할 때, 이 자료의 대푯값으로 적절한 것을 구하고, 그 이유를 설명하시오.

(단위: mm)

255	245	245	265	270	255	270
270	260	270	270	275	270	260

공학적 도구를 이용하여 대푯값을 어떻게 구할까?

자료 수집

다음은 어느 해 17개 시도별 무형 문화재 등록 수를 조사한 자료이다.

(단위: 건)

49	24	18	28	20	24	5	3	67
29	27	53	83	50	36	39	21	

(출처: 통계청, 2018)

공학적 도구를 이용하여 위의 자료에 대한 대푯값을 구해 보자.

자료 입력

이 단원에서 이용한 공학적 도구는 '이지통계(http://www.ebsmath.co.kr/easyTong/etMiddle)'이다.

❶ 위의 자료의 값을 오른쪽 그림과 같이 세로 칸에 한 열로 입력한다.

자료 입력 입력도구 통겟값 설정

No.	자료
1	49
2	24
3	18
4	28
5	20
6	24
7	5
8	3

자료 정리

❷ [통곗값]을 누르면 평균, 중앙값, 최빈값이 나타나고, 원하는 대푯값을 눌러 그 값을 구한다.

자료 입력 입력도구 통겟값 설정

평균 : 33.8824
중앙값 : 28
최빈값 : 24
최댓값 :
최솟값 :

No.	자료
1	49
2	24
3	18
4	28
5	20
6	24
7	5
8	3

| 참고 | '통그라미(http://tong.kostat.go.kr)'를 이용하여 대푯값을 구할 수도 있다.

우리 반 학생들의 작년 한 해 동안의 봉사 활동 시간을 조사하고, 다음 물음에 답하시오.

(1) 조사한 자료에 대한 대푯값으로 적절한 것이 무엇인지 토론하시오.

(2) 공학적 도구를 이용하여 조사한 자료의 대푯값을 구하시오.

산포도

• 분산과 표준편차의 의미를 이해하고, 이를 구할 수 있다.

분산과 표준편차를 어떻게 구할까?

개 념 열 기

다음 표는 어느 해 전국 티볼 대회에 출전한 두 학교 A, B가 각각 10번의 예선 경기에서 얻은 점수를 조사하여 나타낸 것이다.

(단위: 점)

학교	점수									
A	4	5	5	5	6	5	4	5	6	5
B	3	4	3	5	4	5	6	7	6	7

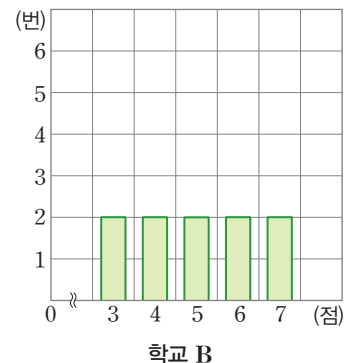
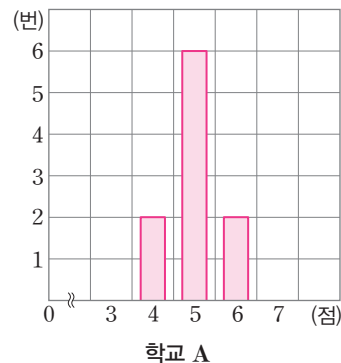
1 두 학교 A, B의 점수의 평균을 각각 구하시오.

2 두 학교 A, B의 점수를 각각 막대그래프로 그리고, 그 분포 상태를 비교하시오.

위의 개념 열기에서 두 학교 A, B의 점수의 평균은 5점으로 서로 같지만 두 학교 A, B의 점수를 각각 막대그래프로 나타내면 다음과 같이 분포 상태가 서로 다름을 알 수 있다.

즉, 학교 A의 점수 분포는 평균 5점을 중심으로 가까이 모여 있지만, 학교 B의 점수 분포는 평균 5점을 중심으로 좌우로 넓게 흩어져 있다.

학교 A

학교 B

이와 같이 두 자료의 평균은 같아도 흩어져 있는 정도는 서로 다를 수 있다.

따라서 자료의 분포 상태의 특성을 알아보기 위해서는 대푯값 이외에도 자료가 흩어져 있는 정도를 살펴보아야 한다.

자료가 흩어져 있는 정도를 하나의 수로 나타낸 값을 **산포도**라고 한다.

산포도는 자료가 평균을 중심으로 흩어져 있는 정도를 이용하여 구할 수 있다.

편차의 절댓값이 클수록 그 자료의 값은 평균에서 멀리 떨어져 있고, 편차의 절댓값이 작을수록 그 자료의 값은 평균에 가까이 있다.

어떤 자료가 있을 때, 각 자료의 값에서 평균을 뺀 값을 **편차**라고 한다. 즉,

$$(\text{편차})=(\text{자료의 값})-(\text{평균})$$

이다.

앞의 개념 열기에서 두 학교 A, B의 점수의 편차와 그 편차의 합을 각각 구하여 표로 나타내면 다음과 같다.

(단위: 점)

학교	편차										합
A	−1	0	0	0	1	0	−1	0	1	0	0
B	−2	−1	−2	0	−1	0	1	2	1	2	0

이와 같이 편차의 합은 항상 0이므로 편차의 평균도 0이 되어 편차의 평균으로는 자료가 평균을 중심으로 흩어져 있는 정도를 알 수 없다.

이때 편차를 제곱하면 모두 음이 아닌 수가 되므로 편차의 제곱의 평균을 이용하면 산포도를 구할 수 있다.

어떤 자료의 편차의 제곱의 평균을 그 자료의 **분산**이라 하고, 분산의 음이 아닌 제곱근을 **표준편차**라고 한다.

자료의 분산 또는 표준편차가 클수록 각 자료의 값이 평균을 중심으로 좌우로 넓게 흩어져 있고, 분산 또는 표준편차가 작을수록 각 자료의 값이 평균을 중심으로 모여 있다.

피어슨
(Pearson, K., 1857~1936)
영국의 수학자. 표준편차라는 용어를 처음 사용하였다.

이상을 정리하면 다음과 같다.

분산과 표준편차

- $(\text{분산})=\dfrac{\{(\text{편차})^2\text{의 합}\}}{(\text{자료의 개수})}$
- $(\text{표준편차})=\sqrt{(\text{분산})}$

다음은 학생 5명이 지난 한 달 동안 대중교통을 이용한 횟수를 조사한 자료이다. 이 자료의 분산과 표준편차를 각각 구하시오.

(단위: 회)

6	5	8	2	4

$(\text{평균})=\frac{6+5+8+2+4}{5}=\frac{25}{5}=5(\text{회})$

$(\text{분산})=\frac{(6-5)^2+(5-5)^2+(8-5)^2+(2-5)^2+(4-5)^2}{5}=\frac{20}{5}=4$

$(\text{표준편차})=\sqrt{4}=2(\text{회})$

답 분산: 4, 표준편차: 2회

다음 표는 재연이와 세원이가 각각 5번의 볼링 경기에서 얻은 점수를 조사하여 나타낸 것이다.

(단위: 점)

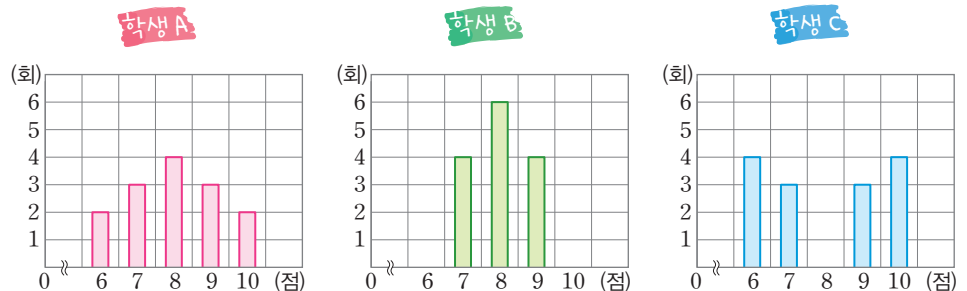
재연	120	126	123	124	122
세원	92	96	93	98	96

(1) 재연이와 세원이의 볼링 점수의 평균과 표준편차를 각각 구하시오.

(2) 재연이와 세원이 중에서 볼링 점수의 변화가 더 작은 사람을 말하고, 그 이유를 설명하시오.

문제를 해결하고 설명할 때는

- 수학적 표현의 의미를 이해한다.
- 자신의 의견을 논리적으로 설명한다.

다음 막대그래프는 세 학생 A, B, C가 각각 과녁판에 14회씩 활을 쏘아 얻은 점수를 조사하여 나타낸 것이다. 세 학생 중에서 점수의 분산이 가장 작은 학생을 말하고, 그 이유를 설명하시오.

◆ 공학적 도구를 이용하여 분산과 표준편차를 어떻게 구할까?

자료 수집

다음은 어느 해 8월 두 지역 A, B에 비가 온 날의 강수량을 조사한 자료이다.

지역 A (단위: mm)

2	5	2.5	5.5
0.5	1.5	93.5	2
0.5	124.5	5	31.5
14	9		

지역 B (단위: mm)

0.5	4.8	3.1	2.8
15.8	12.3	34.9	4.4
0.9	4.7	9.1	89.1
10.2	0.4	80.7	6.7

(출처: 날씨누리, 2017)

공학적 도구를 이용하여 위의 두 자료의 분산과 표준편차를 각각 구해 보자.

자료 입력

❶ [설정]을 눌러 [두 자료]를 선택한다. 또 [모아보기]를 누른 후, 두 자료의 값을 각각 세로 칸에 한 열로 입력한다.

No.	자료 (지역 A)	자료 (지역 B)
1	2	0.5
2	5	4.8
3	2.5	3.1
4	5.5	2.8
5	0.5	15.8

자료 정리

❷ [통곗값]을 누른 후, 하단의 화살표를 누르면 평균, 분산, 표준편차가 나타나고, 각각을 눌러 그 값을 구한다. 각 자료의 탭을 눌러 평균, 분산, 표준편차를 각각 구하여 비교한다.

에어코리아 환경부 대기환경정보(http://www.airkorea.or.kr)에서 내가 살고 있는 지역과 다른 한 지역의 작년 한 해 동안 월별 미세 먼지 오염도를 조사하고, 다음 물음에 답하시오.

(1) 공학적 도구를 이용하여 두 지역의 작년 한 해 동안 월별 미세 먼지 오염도의 평균, 분산, 표준편차를 각각 구하시오.

(2) 두 지역 중에서 월별 미세 먼지 오염도의 변화가 더 작은 지역을 말하시오.

미세 먼지 오염도는 에어코리아 환경부 대기환경 정보 누리집의 통계 정보 ➔ 대기환경 연월보에서 확인할 수 있다.

중단원 학습 점검

개념 정리

- **대푯값**
  자료 전체의 중심 경향이나 특징을 대표적으로 나타내는 값
- **중앙값**
  자료를 작은 값부터 크기순으로 나열할 때, 자료의 중앙에 위치한 값
  ❶ 자료의 개수가 홀수이면 ➔ 중앙에 위치한 하나의 값
  ❷ 자료의 개수가 짝수이면 ➔ 중앙에 위치한 두 값의 평균
- **최빈값**
  자료의 값 중에서 가장 많이 나타난 값
- **산포도**
  자료가 흩어져 있는 정도를 하나의 수로 나타낸 값
- **편차**
  (편차)=(자료의 값)−(평균)
- **분산과 표준편차**
  $(\text{분산})=\dfrac{\{(\text{편차})^2\text{의 합}\}}{(\text{자료의 개수})}$, $(\text{표준편차})=\sqrt{(\text{분산})}$

다음 문장이 옳으면 O, 옳지 않으면 X를 ( ) 안에 쓰시오.

1 자료 1, 2, 3, 4, 5, 6의 중앙값은 3과 4이다. ( )

2 자료 1, 1, 2, 3, 4의 최빈값은 1이다. ( )

3 편차의 합은 0이다. ( )

4 어떤 자료의 편차의 평균을 그 자료의 분산이라고 한다. ( )

5 분산의 음이 아닌 제곱근을 표준편차라고 한다. ( )

기초 문제

1 다음 줄기와 잎 그림은 소윤이네 반 학생 20명의 하루 동안 이모티콘 사용 건수를 조사하여 그린 것이다. 이 자료의 중앙값과 최빈값을 각각 구하시오.

이모티콘 사용 건수 (1|0은 10건)

줄기	잎
1	0 1 2 3
2	1 4 6 6 8 9
3	2 3 6 8 9
4	0 3 5
5	3 7

2 다음은 지호네 모둠 학생 5명이 일주일 동안 취미 활동을 한 시간을 조사한 자료이다.

(단위: 시간)

2	5	6	3	4

(1) 이 자료의 평균을 구하시오.

(2) 다음 표의 빈칸을 알맞게 채우시오.

자료(시간)	2	5	6	3	4
편차(시간)					
$(\text{편차})^2$					

(3) 이 자료의 분산과 표준편차를 각각 구하시오.

정답 및 해설 ▸ 268쪽

기본 문제

3 다음은 윤아네 가족의 어느 해의 월별 전기 사용량을 조사한 자료이다. 이 자료의 대푯값으로 적절한 것을 구하고, 그 이유를 설명하시오.

(단위: kWh)

195	210	200	230	215	220
280	1080	210	205	198	190

4 다음은 5개의 자료를 작은 값부터 크기순으로 나열한 것이다. 이 자료의 평균과 중앙값이 모두 같을 때, x의 값을 구하시오.

2	4	5	7	x

5 다음 표는 어느 야구팀이 6일 동안 경기에서 친 안타 수의 편차를 나타낸 것이다. 안타 수의 평균이 10개일 때, 물음에 답하시오.

요일	월	화	수	목	금	토
편차(개)	2	−2	a	1	0	3

(1) a의 값을 구하시오.
(2) 수요일에 친 안타 수를 구하시오.

6 다음은 학생 5명의 기술 수행 평가 점수의 편차이다. 이 학생들의 기술 수행 평가 점수의 표준편차를 구하시오.

(단위: 점)

−3	0	4	−2	1

7 다음 표는 서준이와 준우가 6회에 걸쳐 실시한 턱걸이 개수를 조사하여 나타낸 것이다. 두 학생의 턱걸이 개수의 평균이 15번으로 같을 때, 물음에 답하시오.

(단위: 번)

서준	14	14	15	14	16	17
준우	13	11	15	17	18	16

(1) 두 학생의 턱걸이 개수의 표준편차를 각각 구하시오.
(2) 두 학생 중에서 턱걸이 개수가 더 고르게 나타난 학생을 말하고, 그 이유를 설명하시오.

도전 문제

8 다음을 읽고, 두 사람의 대화를 모두 만족시키는 a, b의 값을 각각 구하시오.

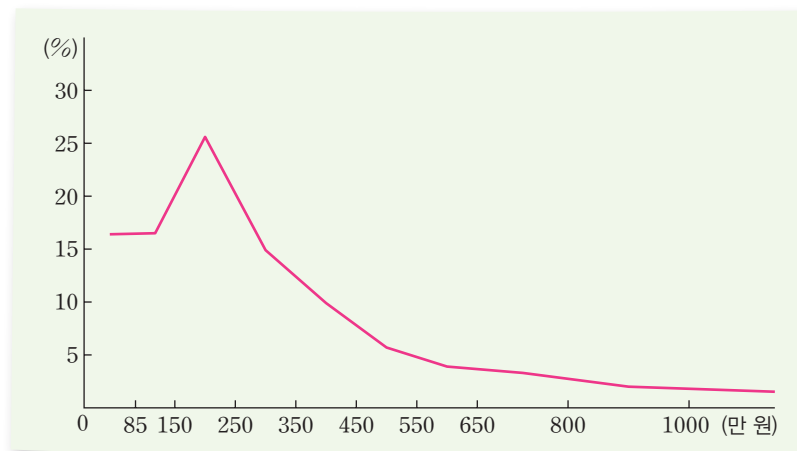
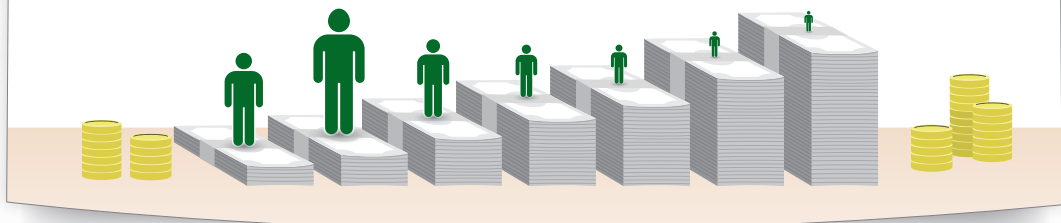
스스로 푸는 자기 주도 학습 수학 익힘책 ▸ 240쪽

수행 과제 상대도수의 분포 그래프에서 평균과 중앙값을 비교해 볼까?

의사소통

활동 목표 중앙값과 평균의 의미를 이해하고, 이를 이용하여 자료를 분석할 수 있다.

다음 글을 읽고, 물음에 답해 보자.

다음 그림은 통계청에서 발표한 우리나라 임금 근로자 약 15만 명의 소득 분포 자료를 바탕으로 상대도수의 분포를 나타낸 것이다. 이때 이 자료에서 월 소득의 평균은 281만 원, 중앙값은 209만 원이다.

(출처: 통계청, 2017)

위의 그림과 같이 오른쪽으로 꼬리가 긴 모양의 소득 분포에서는 월 소득의 평균이 중앙값보다 크다고 한다. 그 이유를 설명해 보자.

적절한 대푯값의 선택

미국 프로 야구에서는 한때 선수들의 파업으로 경기가 장기간 취소된 적이 있다. 파업 이유는 선수들과 구단주들 사이의 마찰 때문이다. 구단주들은 여론을 유리한 방향으로 이끌기 위해 “평균 연봉이 13억 원이나 되는 선수들이 파업 하다니….”라고 하면서 엄청난 연봉을 받는 선수들이 욕심을 부린다는 식으로 홍보해 야구 팬들의 비난을 유도하였다. 이러한 의도는 적중하였고 한 아동 팬은 “돈을 더 원하면 내 용돈을 가져가라.”라고 쓴 팻말을 들고 야구장에서 항의 표시를 하였다고 한다.

실제로 그 당시 선수들의 평균 연봉은 13억여 원이었다. 그러나 중앙값은 그보다 훨씬 적은 4억여 원, 최빈값은 3억여 원이었다. 이와 같은 차이가 발생한 것은 고액 연봉을 받는 선수들이 몇 명 있었기 때문이다.

따라서 구단주들의 입장에서는 여론을 유리한 방향으로 이끌기 위해 평균 연봉을 이용하는 것이 당연했고, 이에 대해 선수들의 입장에서는 최빈값을 이용하여 대응했어야 했다. 반면에 야구 팬들의 입장에서는 중앙값을 고려하여 양측 주장의 적절성을 판단했어야 했다.

(참고 자료: 통계청, “손안의 통계: 당신의 답답함을 해결해주는 9가지 처방”)

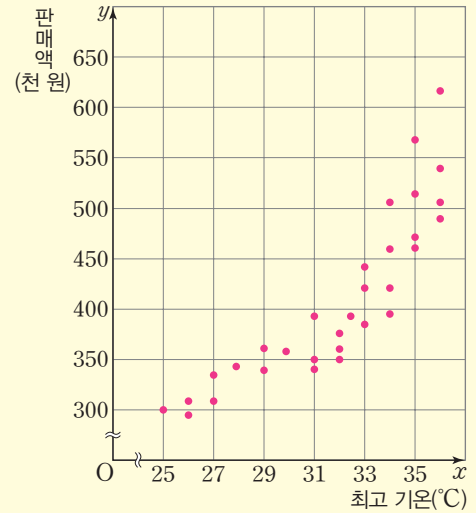
2 상관관계

수학 + 경영

오늘날 날씨 정보는 재해 예방 수단의 차원을 넘어 부가 가치를 창출하는 중요한 요소로 인식되고 있다. 기업에서는 날씨에 따른 소비자의 구매 욕구, 선호 상품 등을 분석하여 마케팅 전략을 세우기도 한다. 특히 빙과류, 음료수, 냉난방기 등과 같은 계절상품 생산 업체에서는 기상 예보를 모니터링하여 날씨로부터 발생되는 위험 요소를 최소화하고 있다.

(참고 자료: 한국기상산업진흥원, “날씨를 알면 ‘경영도 맑음’(기상정보 활용 성공사례집)”)

단원 활동 일별 최고 기온과 어느 편의점의 빙과류 판매액에 대한 그림을 보고, 두 변량 사이의 관계를 알아보자.

오른쪽 그림은 어느 해 8월 한 달 동안의 일별 (최고 기온, 판매액)을 좌표로 하는 점을 좌표평면 위에 나타낸 것이야.

활동 1 위의 그림에서 점들의 전체적인 경향이 대체로 오른쪽 위로 향하는지 오른쪽 아래로 향하는지 말해 보자.

활동 2 위의 그림에서 최고 기온이 높을수록 빙과류 판매액이 대체로 크다고 할 수 있는지 말해 보자.

두 변량 사이의 관계를 알 수 있는 방법을 알아볼까?

위의 활동으로 **알게 된 것**과 **나의 학습 계획**을 적어 보자.

알게 된 것 ▶ 두 변량 사이의 관계를 나타낸 그림의 특징을 말할 수 있다. 예 ☐ 아니요 ☐

학습할 내용 ▶ 산점도 ▶ 상관관계

학습 계획

학습 계획안 예시

- 예습과 복습을 열심히 하겠다.
- 수업 시간에 집중하겠다.
- 수학에 대한 자신감을 키우겠다.
- 모둠 활동에 적극적으로 참여하겠다.

산점도와 상관관계

• 자료를 산점도로 나타내고, 이를 이용하여 상관관계를 말할 수 있다.

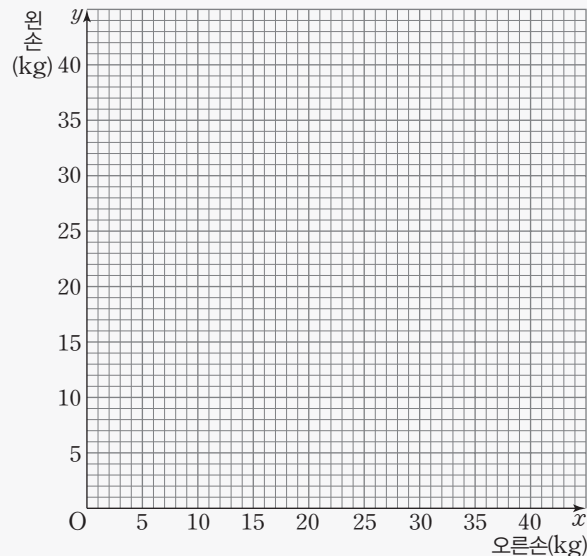
산점도와 상관관계는 무엇일까?

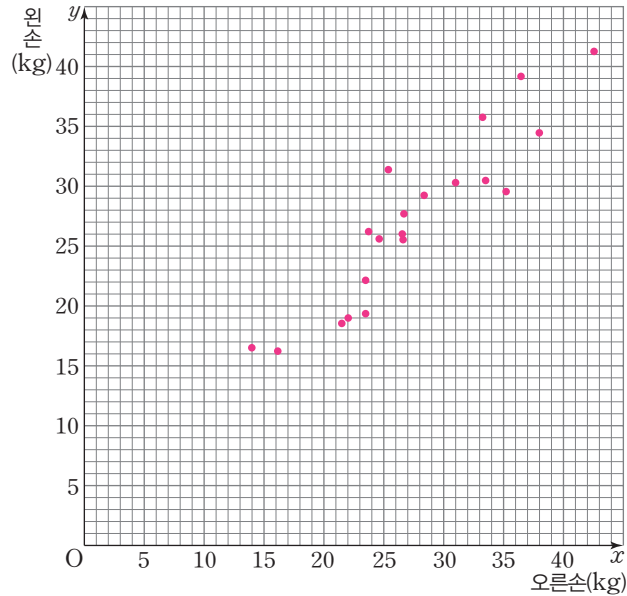
개 념 열 기

악력
손으로 물건을 쥐는 힘을 말한다. 악력계를 사용하여 측정할 수 있고, 단위는 kg으로 표시한다.

다음 표는 어느 중학교 3학년 학생 20명의 양손의 악력을 조사하여 나타낸 것이다.

번호	오른손(kg)	왼손(kg)	번호	오른손(kg)	왼손(kg)
1	36.3	39.1	11	14.0	16.5
2	23.3	22.1	12	26.7	26.0
3	26.8	25.4	13	26.9	27.9
4	24.9	25.6	14	28.3	29.2
5	33.6	30.4	15	21.5	18.6
6	35.2	29.7	16	16.1	16.2
7	42.7	41.2	17	38.0	34.6
8	23.4	19.4	18	23.9	26.1
9	25.3	31.3	19	22.0	19.0
10	33.2	35.9	20	31.0	30.3

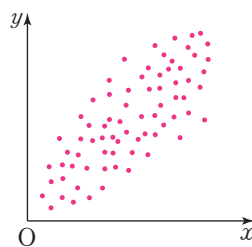
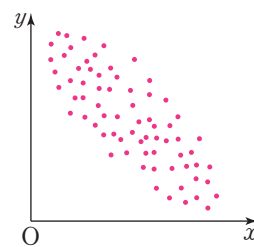
1 위의 자료에서 학생들의 오른손의 악력을 x kg, 왼손의 악력을 y kg이라고 할 때, 순서쌍 (x, y)를 좌표로 하는 점을 아래의 좌표평면 위에 나타내시오.

2 오른손의 악력이 셀수록 왼손의 악력도 대체로 세다고 할 수 있는지 말하시오.

앞의 개념 열기에서 학생들의 오른손의 악력 x kg, 왼손의 악력 y kg에 대한 순서쌍 (x, y)를 좌표로 하는 점을 좌표평면 위에 나타내면 다음과 같다.

이와 같이 두 변량 x, y의 순서쌍 (x, y)를 좌표평면 위에 나타낸 그림을 두 변량 x, y의 **산점도**라고 한다.

위의 산점도를 보면 x의 값이 증가함에 따라 y의 값도 대체로 증가하는 경향이 있다고 할 수 있다. 즉, 오른손의 악력이 셀수록 왼손의 악력도 대체로 세다고 할 수 있다.

일반적으로 두 변량에 대하여 한 변량의 값이 변함에 따라 다른 변량의 값이 변하는 경향이 있을 때, 이 두 변량 사이의 관계를 **상관관계**라고 한다.

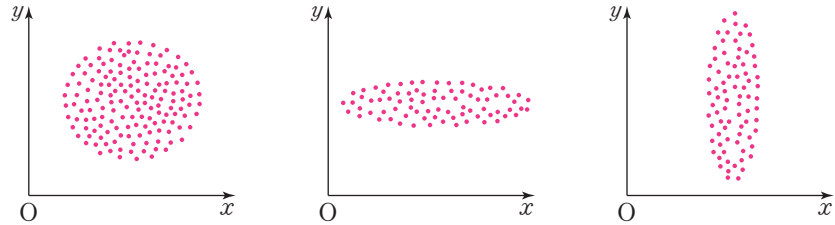
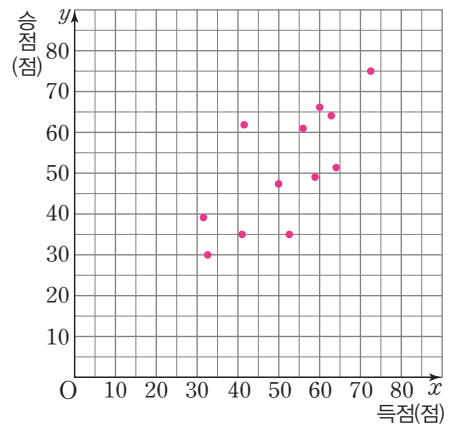
특히 한 변량의 값이 증가함에 따라 다른 변량의 값도 대체로 증가하는 경향이 있을 때, 이 두 변량 사이에는 양의 상관관계가 있다고 한다. 거꾸로 한 변량의 값이 증가함에 따라 다른 변량의 값이 대체로 감소하는 경향이 있을 때, 이 두 변량 사이에는 음의 상관관계가 있다고 한다.

산점도에서 점들이 오른쪽 위로 향하는 경향이 있으면 양의 상관관계가 있고, 오른쪽 아래로 향하는 경향이 있으면 음의 상관관계가 있다.

양의 상관관계　　음의 상관관계

한편 양의 상관관계도 없고 음의 상관관계도 없는 경우에는 두 변량 사이에 상관관계가 없다고 한다.

다음 그림은 상관관계가 없는 경우이다.

산점도에서 점들이 오른쪽 위로 향하거나 오른쪽 아래로 향하는 경향이 있지 않을 때, 상관관계는 없다.

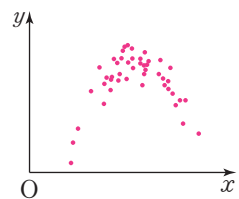
오른쪽 그림은 어느 해 12개 축구팀의 득점과 승점에 대한 산점도이다. 두 변량 사이의 상관관계를 말하시오.

축구에서 득점은 경기에서 얻은 점수이고, 승점은 경기의 승패에 따라 계산한 점수이다.

의사소통

토의할 때는

- 자신의 생각을 명확하게 전달한다.
- 타인의 의견을 주의 깊게 듣는다.

오른쪽 그림은 두 변량 x, y에 대한 산점도이다. 이 산점도에 대한 준우와 하은이의 의견 중에서 누구의 의견이 옳은지 토의하시오.

준우

x의 값이 증가함에 따라 y의 값이 대체로 감소하는 경향이 있으니까, 두 변량 사이에는 음의 상관관계가 있어.

하은

글쎄, x의 값이 증가함에 따라 y의 값이 대체로 증가한다고 할 수도 없고, 감소한다고 할 수도 없으니까 두 변량 사이에는 상관관계가 없지.

정보 처리

◆ 공학적 도구를 이용하여 산점도를 어떻게 그릴까?

자료 수집

다음 표는 어느 공원에서 건강 달리기를 한 20명의 운동 시간과 운동 후 1분 동안의 심박수를 조사하여 나타낸 것이다.

운동 시간(분)	40	30	60	70	30	50	70	30	80	90
심박수(회)	83	79	88	89	81	87	90	80	90	93
운동 시간(분)	90	45	60	40	50	100	45	80	30	60
심박수(회)	92	86	89	85	88	95	88	91	82	90

공학적 도구를 이용하여 운동 시간과 심박수에 대한 산점도를 그려 보자.

자료 입력

❶ [설정]을 눌러 [두 자료]를 선택한다. 또 [모아보기]를 누른 후, 두 자료의 값을 각각 세로 칸에 한 열로 입력한다.

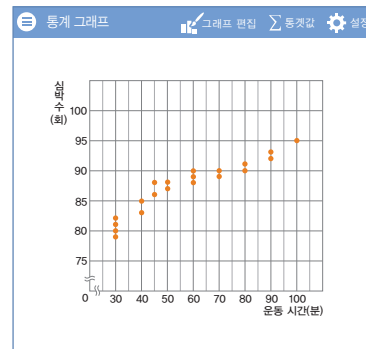
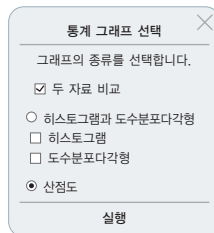
자료 입력 · 입력도구 · 통곗값 · 설정
☑ 두 자료 · 운동 시간 / 심박수 · 모아보기 · 배경색

No.	자료 (운동 시간)	자료 (심박수)
1	40	83
2	30	79
3	60	88
4	70	89
5	30	81

자료 정리

❷ [통계 그래프]를 누른 후 [두 자료 비교]와 [산점도]를 선택하고 실행하면 두 변량에 대한 산점도가 그려진다.

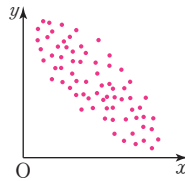
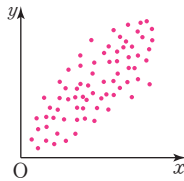
날씨누리(http://www.weather.go.kr)에서 내가 살고 있는 지역의 작년 11월 한 달 동안의 일조 시간과 운량을 조사하고, 다음 물음에 답하시오.

(1) 공학적 도구를 이용하여 조사한 일조 시간과 운량에 대한 산점도를 그리시오.

(2) 조사한 두 변량 사이의 상관관계를 말하시오.

일조 시간과 운량은 날씨누리 누리집의 관측 자료 ➔ 지상 관측 자료 ➔ 과거 자료 ➔ 요소별 자료 에서 확인할 수 있다.

중단원 학습 점검

개념 정리

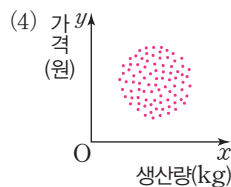
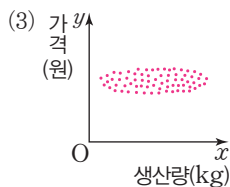
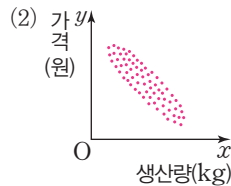
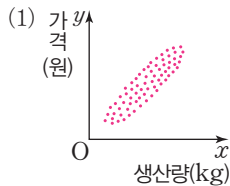
- **산점도**
  두 변량 x, y의 순서쌍 (x, y)를 좌표평면 위에 나타낸 그림
- **상관관계**
  두 변량에 대하여 한 변량의 값이 변함에 따라 다른 변량의 값이 변하는 경향이 있는 관계
- **양의 상관관계**
  두 변량에 대하여 한 변량의 값이 증가함에 따라 다른 변량의 값도 대체로 증가하는 경향이 있는 관계
- **음의 상관관계**
  두 변량에 대하여 한 변량의 값이 증가함에 따라 다른 변량의 값이 대체로 감소하는 경향이 있는 관계

O, X 문제

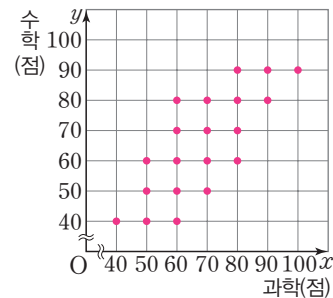
다음 문장이 옳으면 O, 옳지 않으면 X를 (　) 안에 쓰시오.

1 하루 동안의 낮의 길이와 밤의 길이 사이에는 양의 상관관계가 있다. (　)

2 겨울철에 기온이 낮아짐에 따라 손난로의 판매량이 대체로 증가한다고 할 때, 기온과 손난로의 판매량 사이에는 음의 상관관계가 있다. (　)

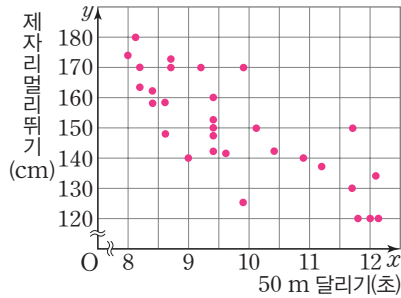
기초 문제

1 한 해 동안 생산된 배추의 양이 많을수록 그 해의 배추 가격이 떨어진다고 한다. 그 해의 배추 생산량을 x kg, 배추 가격을 y원이라고 할 때, 다음 중에서 x와 y 사이의 상관관계를 나타낸 산점도로 알맞은 것을 찾으시오.

(1)
(2)
(3)
(4)
기본 문제

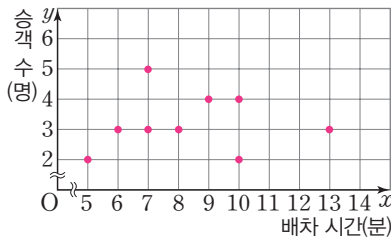
2 다음 그림은 학생 20명의 과학 성적과 수학 성적에 대한 산점도이다.

(1) 과학 성적이 가장 낮은 학생과 가장 높은 학생의 수학 성적을 비교하시오.

(2) 과학 성적과 수학 성적 사이의 상관관계를 말하시오.

정답 및 해설 ▸ 269쪽

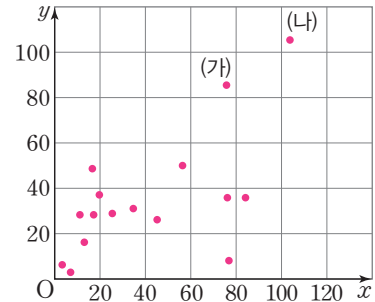
3 다음 그림은 학생 30명의 50 m 달리기 기록과 제자리멀리뛰기 기록에 대한 산점도이다.

(1) 50 m 달리기 기록과 제자리멀리뛰기 기록 사이의 상관관계를 말하시오.

(2) 50 m 달리기를 잘하는 학생은 제자리멀리뛰기도 대체로 잘한다고 할 수 있는지 말하시오.

4 다음 그림은 어느 정거장에서 정차하는 버스 9대의 배차 시간과 각 버스를 기다리는 승객 수에 대한 산점도이다.

(1) 배차 시간이 가장 짧은 버스를 기다리는 승객 수와 가장 긴 버스를 기다리는 승객 수를 각각 구하시오.

(2) 버스의 배차 시간과 버스를 기다리는 승객 수 사이의 상관관계를 말하시오.

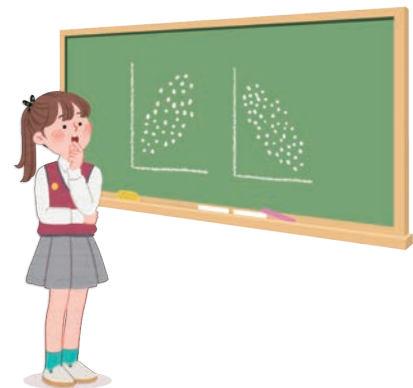
도전 문제

5 다음 그림은 두 변량 x와 y에 대한 산점도이다.

(1) 두 변량 x와 y 사이의 상관관계를 말하시오.

(2) 위의 산점도에서 두 점 (가), (나)를 지웠을 때, 두 변량 x와 y 사이의 상관관계를 말하시오.

(3) 위의 산점도에 다음 5개의 자료를 추가하였을 때, 두 변량 x와 y 사이의 상관관계를 말하시오.

x	60	70	80	100	110
y	60	80	80	100	100

스스로 푸는 자기 주도 학습 수학 익힘책 ▸ 241쪽

수행 과제 연령과 최대 심박수 사이에는 어떤 상관관계가 있을까? 문제 해결

활동 목표 수집한 자료를 산점도로 나타내고, 이를 이용하여 상관관계를 말할 수 있다.

최대 심박수에 대한 다음 글을 읽고, 문제를 해결해 보자.

심박수는 단위 시간당 심장 박동의 횟수이다. 건강한 사람의 평상시 심박수는 분당 60~100회 정도인데 운동의 강도가 세지면 심박수는 빨라진다. 그러나 운동의 강도가 아무리 세도 심박수는 일정 수준 이상은 넘지 않는데 이 수준을 최대 심박수라고 한다. 최대 심박수는 연령, 체력 등에 따라 다르다. 10세 미만의 어린이들에게 연령과 최대 심박수 사이의 상관관계는 거의 없지만 15세 이상의 사람들에게는 상관관계가 있다고 한다.

최대 심박수는 20~30분 동안 달리기 또는 수영 등의 운동을 충분히 한 후, 스마트폰 앱을 이용하여 측정할 수 있다.

(참고 자료: Hottenrott, K., “Training with the Heart Rate Monitor”)

1 10대, 20대, …, 60대의 연령대별로 각각 3~4명의 건강한 사람을 선정하여 연령과 분당 최대 심박수를 조사해 보자.

2 1에서 수집한 자료에 대한 산점도를 그리고, 연령과 최대 심박수 사이에 어떤 상관관계가 있는지 알아보자.

얼음의 온도와 마찰력의 상관관계

빙상 경기에서는 얼음의 마찰력이 승패를 가르는 중요한 요인이다. 봅슬레이에서 썰매에 장착하는 스케이트 날을 뜨겁게 달군 팀이 국제 대회에서 우승한 이후로, 스케이트 날의 온도에 대한 규정이 새로 만들어졌다. 뜨거운 스케이트 날의 온도는 얼음의 온도를 변화시켜 마찰력을 줄일 수 있기 때문이다. 이와 같이 얼음의 온도와 마찰력은 빙상 경기에 큰 영향을 미친다.

컬링 또한 얼음의 상태에 따라 스톤의 활주 방향과 속력, 거리 등이 예민하게 바뀔 수 있다. 얼음판을 문지르는 행위인 스위핑(sweeping)은 얼음 표면의 온도를 높여 스톤과 얼음 사이의 마찰력을 줄임으로써 스톤의 속력과 이동 경로를 정하는 데 중요한 역할을 한다.

(참고 자료: 김태완 · 이상철 · 문영진 · 길세기 · 송주호, “2018 컬링 골드 프로젝트”)

대단원 학습 평가

1 ●○○

다음 줄기와 잎 그림은 20개 나라의 ICT(정보통신기술) 이용 부문 점수를 조사하여 그린 것이다. ICT 이용 부문 점수의 중앙값과 최빈값을 각각 구하시오.

ICT 이용 부문 점수 (6|5는 6.5점)

줄기	잎
6	5 6 7 9 9
7	3 6 6 7 8 9
8	0 1 2 3 3 4 4 4 8

(출처: IT STAT, 2016)

2 ●○○

다음은 세민이네 모둠 학생 9명이 1년 동안 여행한 횟수를 조사한 자료이다. 이 자료의 최빈값이 8회일 때, 중앙값을 구하시오.

(단위: 회)

7　10　11　8　6　7　x　8　5

3 ●●○

다음은 어느 동호회 회원 5명의 나이에 대한 설명이다. 이때 회원 5명의 나이의 중앙값을 구하시오.

- 나이의 최빈값은 18세이다.
- 회원 중에서 한 사람의 나이는 15세이다.
- 나이가 가장 적은 회원은 9세이다.
- 회원 5명의 평균 나이는 14.8세이다.

4 ●●○

다음 설명 중에서 옳은 것은?

① 분산이 클수록 표준편차는 작다.
② 표준편차는 편차의 제곱의 평균이다.
③ 산포도에는 평균, 분산, 표준편차가 있다.
④ 편차의 절댓값이 작을수록 그 자료의 값은 평균에 가깝다.
⑤ 자료 전체의 중심 경향이나 특징을 하나의 수로 나타낸 값을 산포도라고 한다.

5 ●●○

다음 표는 5개 도시의 같은 시각 기온의 편차를 나타낸 것이다. 5개 도시의 기온의 평균이 19 ℃일 때, 물음에 답하시오.

도시	서울	대구	대전	부산	광주
편차(℃)	1	-2	0	2	-1

(1) 기온이 평균 기온보다 높은 도시를 모두 말하시오.
(2) 광주의 기온을 구하시오.

6 ●●○

다음은 8개 자료의 편차이다. 수 x의 값을 구하고, 이 자료의 표준편차를 구하시오.

-1　2　4　-3　x　-2　-1　2

정답 및 해설 ▸ 270쪽

●●○

7 다음 표는 어느 중학교 3학년 세 학급의 학생들이 1년 동안 영화를 관람한 횟수의 평균과 표준편차를 나타낸 것이다. 보기 중에서 옳은 것을 모두 찾으시오.

(단위: 회)

반	1	2	3
평균	5	8	8
표준편차	2	1	3

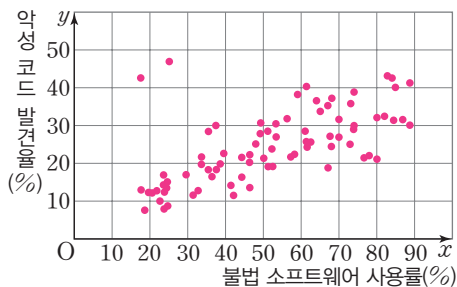
• 보기 •

ㄱ. 2반이 1반보다 평균적으로 영화 관람을 더 많이 했다.

ㄴ. 영화를 가장 많이 본 학생은 1반에 있다.

ㄷ. 영화 관람 횟수는 2반이 3반보다 더 고르다.

●●○

8 다음 그림은 80명의 불법 소프트웨어 사용률과 악성 코드 발견율에 대한 산점도이다. 불법 소프트웨어 사용률과 악성 코드 발견율 사이의 상관관계를 말하시오.

서 술 형 문 제

[9~12] 다음 문제의 풀이 과정을 자세히 쓰시오.

●●○

9 다음은 9개의 자료를 작은 값부터 크기순으로 나열한 것이다. 이 자료의 평균, 중앙값, 최빈값이 모두 같을 때, a, b의 값을 각각 구하시오.

4　6　6　7　8　8　a　11　b

●●○

10 다음 표는 어느 골프 대회에 참가한 두 선수 A, B의 4라운드 동안의 타수를 조사하여 나타낸 것이다.

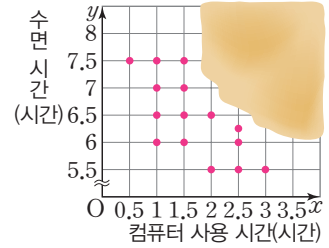
(단위: 타)

A	68	74	72	70
B	71	75	67	71

(1) 두 선수 A, B의 타수의 평균과 분산을 각각 구하시오.

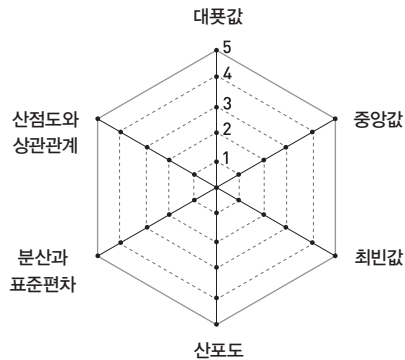
(2) 두 선수 A, B 중에서 어느 선수의 타수가 더 고른지 설명하시오.

11 4개의 자료 x, y, 6, 2의 평균이 4이고, 분산이 12일 때, 4개의 자료 x, y, 3, 5의 표준편차를 구하시오.

12 오른쪽 그림은 윤아네 반 학생 21명의 하루 동안 컴퓨터 사용 시간과 수면 시간에 대한 산점도의 일부이다. 얼룩진 부분의 자료가 다음과 같을 때, 컴퓨터 사용 시간과 수면 시간 사이의 상관관계를 말하시오.

(단위: 시간)

컴퓨터 사용 시간	2	3	2.5	3.5	3.5	3
수면 시간	8	7.5	7.5	7	7.5	8

자기 평가

1 이 단원에서 학습한 내용에 대한 나의 성취 수준을 다음 그림에 점으로 표시하고, 이웃한 점을 선으로 연결해 보자.

|성취 수준|

1수준: 개념을 이해하기 어려웠다.
2수준: 문제를 해결하기 어려웠다.
3수준: 문제를 일부 해결하였다.
4수준: 문제를 대부분 해결하였다.
5수준: 문제를 모두 해결하였다.

이해가 부족한 내용은 본문을 복습! 문제가 더 필요하면 수학 익힘책 ▶ 240~241쪽

2 이 단원을 시작할 때 세운 학습 계획을 잘 실천하였는지 평가해 보고, 이해하기 어려웠던 내용을 적어 보자.

여론 조사 전문가

여론 조사 전문가가 하는 일은?

여론 조사 전문가는 사회적 문제, 공공 정책 개발 및 평가, 계층별 의식 등에 대한 조사 업무를 수행하고 이에 대한 보고서를 작성한다.

여론 조사를 하는 데 수학이 이용되나요?

여론 조사 전문가는 조사를 실시하고 그 결과를 분석할 때, 대중의 의견에 대한 전체적인 특징을 추출하기 위해서 다양한 통계적 분석 기법을 이용한다. 또 분석된 결과를 바탕으로 보고서를 작성하며 통계학의 지식을 비롯하여 경제학, 사회학, 심리학 등의 지식을 활용한다.

수학 + 사회

통계 보고서는 전달하고자 하는 내용을 명확하고 객관적으로 작성하여 읽는 사람이 그 의미를 정확하게 알 수 있어야 한다. 그러나 같은 내용이라도 그것을 어떻게 표현하는가에 따라 읽는 사람이 쉽게 이해할 수 있고 그렇지 않을 수도 있다. 따라서 논리적인 구성과 시각화된 디자인으로 완성도를 높여야 한다.

(참고 자료: 통계청, “손 안의 통계: 당신의 답답함을 해결해주는 9가지 처방”)

친구들아,

스마트폰 사용을 줄이고 운동하자!

작성일: 20○○년 ○○월 ○○일 작성자: 김○○, 박○○, 서○○, 최○○

1 조사 목적

우리나라 청소년의 체격은 좋아졌지만 체력은 떨어졌다는 뉴스를 듣고, 스마트폰 사용 시간이 증가함에 따라 운동 시간이 감소하였다고 추측하였다. 우리의 추측을 확인하기 위해 필요한 자료를 수집하여 분석하고 정리하였다.

2 자료 수집 및 정리, 분석

가. 조사 대상 및 자료 수집 3학년 ○○반 학생 20명의 일주일 동안 스마트폰 사용 시간과 운동 시간

나. 자료 수집 방법 설문지 사용

다. 자료 정리

(1) 스마트폰 사용 시간에 대한 대푯값과 산포도

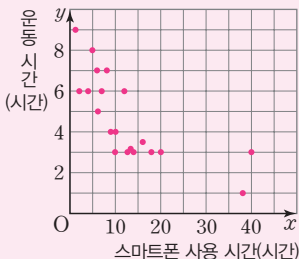
학생 20명의 스마트폰 사용 시간의 평균과 중앙값은 각각 12.6시간, 9.5시간이다. 학생 2명의 스마트폰 사용 시간이 다른 학생들에 비해 길어서 평균이 중앙값보다 크다. 그리고 표준편차는 10.11시간이다.

(2) 운동 시간에 대한 대푯값과 산포도

학생 20명의 운동 시간의 평균과 중앙값은 각각 4.6시간, 4시간이다. 그리고 표준편차는 2시간이다.

(3) 스마트폰 사용 시간과 운동 시간의 비교

학생 20명의 스마트폰 사용 시간과 운동 시간에 대한 대푯값과 산포도를 비교하면 학생들의 스마트폰 사용 시간의 평균이 운동 시간의 평균보다 크고 표준편차도 크다.

스마트폰 사용 시간과 운동 시간에 대한 산점도는 오른쪽 그림과 같고, 스마트폰 사용 시간이 증가할수록 운동 시간이 대체로 감소함을 알 수 있다.

3 결론

우리가 조사한 결과로부터 스마트폰 사용 시간과 운동 시간 사이에 음의 상관관계가 있음을 확인할 수 있다.

통계 보고서 작성 방법

❶ 주제 정하기
❷ 자료 수집 및 방법 정하기
❸ 자료 정리 및 분석하기
❹ 보고서 작성하기

모둠별로 주제를 정해 통계 보고서를 작성해 보자.

활동지 통계 보고서 작성하기

■ 모둠명:

■ 모둠원:

1 서로 관련이 있는 것으로 추측되는 두 자료에 대한 통계 보고서 작성 계획을 세워 보자.

2 **1**에서 작성한 계획서를 바탕으로 수집한 자료를 정리하고 분석하여 통계 보고서를 작성해 보자.

과정 중심 평가	통계 보고서에 알맞은 주제를 정하였는가?
	주제에 적합한 자료를 수집하였는가?
	수집한 자료를 목적에 맞게 정리하고 옳게 분석하였는가?
	대푯값, 산포도, 상관관계 등을 이용하여 통계 보고서를 작성하였는가?

스스로 푸는 자기 주도 학습

수학 익힘책

1. 대푯값과 산포도

정답 및 해설 ▸ 277쪽

●○○

1 다음 중에서 자료의 평균을 대푯값으로 사용하기에 적절하지 않은 것은?

① 3, 4, 5, 6, 7
② 7, 7, 7, 7, 7
③ 2, 4, 6, 8, 10
④ 10, 10, 20, 20, 30
⑤ 10, 10, 10, 10, 500

●●○

2 다음은 6개의 자료를 작은 값부터 크기순으로 나열한 것이다. 이 자료의 중앙값이 11일 때, x의 값을 구하시오.

6	7	x	13	14	57

●○○

3 다음 표는 6개 비타민 음료 100 mL의 열량의 편차를 나타낸 것이다. 비타민 음료의 열량의 평균이 34 kcal일 때, 빈칸에 알맞은 수를 쓰고 음료 C의 열량을 구하시오.

음료	A	B	C	D	E	F
편차(kcal)	-4	16		-16	11	-14

●●○

4 다음은 소윤이네 모둠 학생 8명이 농구 경기에서 자유투 성공 횟수를 조사한 자료이다. 이 자료의 평균과 최빈값이 7회로 같을 때, 분산과 표준편차를 각각 구하시오.

(단위: 회)

4	10	x	7	6	y	5	8

●●○

5 다음 표는 5일 동안 준우와 세원이의 수면 시간을 조사하여 나타낸 것이다. 두 학생의 수면 시간의 표준편차를 각각 구하고, 수면 시간의 변화가 더 큰 학생을 찾아 그 이유를 설명하시오.

(단위: 시간)

준우	6	8	7	8	6
세원	5	9	7	4	10

●●●

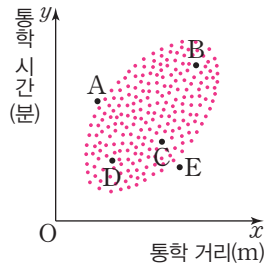
6 5개의 자료 5, 9, 11, a, $5-a$의 분산이 12일 때, a의 값을 모두 구하시오.

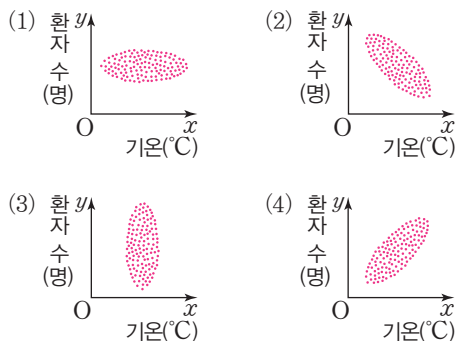
2. 상관관계

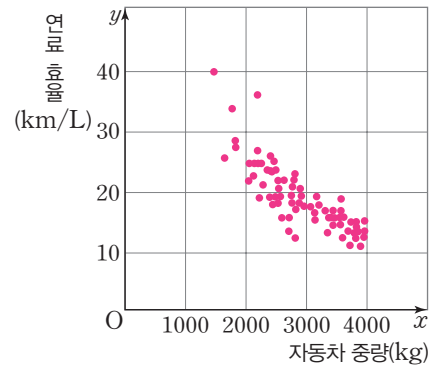
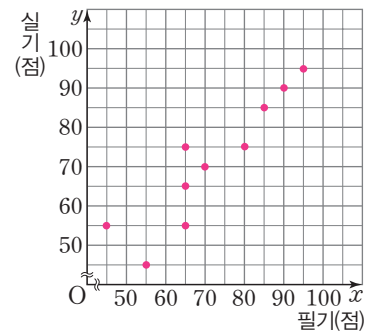
정답 및 해설 ▸ 277쪽

●○○

1 다음 그림은 지호네 학교 학생들의 통학 거리와 통학 시간에 대한 산점도이다. 5명의 학생 A, B, C, D, E 중에서 통학 거리가 짧은데도 통학 시간이 긴 학생을 말하시오.

●○○

2 겨울철 기온이 내려갈수록 감기 환자 수가 증가한다고 한다. 겨울철 기온을 x ℃, 감기 환자 수를 y명이라고 할 때, 다음 중에서 x와 y 사이의 상관관계를 나타낸 산점도로 알맞은 것을 찾으시오.

(1)
(2)
(3)
(4)
●●○

3 다음 그림은 자동차 70대의 중량과 연료 효율에 대한 산점도이다. 자동차의 중량과 연료 효율 사이의 상관관계를 말하시오.

●●○

4 다음 그림은 워드 프로세서 시험에 응시한 학생 10명의 필기 점수와 실기 점수에 대한 산점도이다.

(1) 필기 점수와 실기 점수가 같은 학생은 전체의 몇 %인지 구하시오.

(2) 필기 점수와 실기 점수 사이의 상관관계를 말하시오.